

Note

Utilisation de la numérotation binaire pour le calcul de la consanguinité

J. LEFEBVRE

*I.N.R.A., Laboratoire de Génétique factorielle
Centre de Recherches zootechniques
F - 78350 Jouy-en-Josas*

Résumé

La numérotation binaire des positions des ancêtres dans une généalogie permet de réduire considérablement la place nécessaire en mémoire pour les calculs de consanguinité sur ordinateur. Elle a été utilisée pour réaliser sur micro-ordinateur un programme prenant en considération jusqu'à 16 générations. Sa simplicité et son faible coût de mise en œuvre rendent possibles de longues séries de calculs.

Les programmes de calcul de consanguinité publiés jusqu'à présent demandent des capacités en mémoire et des temps d'exécution, donc entraînent des coûts, limitant sévèrement leurs applications (HAZEL & LUSH, 1950 ; PLUM, 1952, 1954 ; REHFELD *et al.*, 1967 ; ALFONSO PONCE, 1971, 1974 ; LOWE, 1972 ; FALCO, 1975 ; ROSTRON, 1978).

Nous avons indiqué précédemment (LEFEBVRE, 1965) l'intérêt de la numérotation binaire dans la constitution des généalogies mais cette voie ne semble pas avoir été reprise par d'autres auteurs. De notre côté, nous n'avons pas eu la possibilité, jusqu'à ces dernières années, d'utiliser un ordinateur où la manipulation de cette numérotation soit réalisable.

Avec l'arrivée en 1976 d'un micro-ordinateur WANG 2200 la situation a été modifiée mais la mémoire, de 4 K puis de 8 K, était trop restreinte pour envisager des exploitations sur un nombre suffisant de générations.

Lorsque, disposant d'une mémoire de 16 K et de plus de temps de machine, nous avons voulu étudier les instructions binaires nous n'avons trouvé, dans le

chapitre de la brochure consacré par le constructeur à cet aspect, que des indications brèves et hermétiques alors que les autres parties étaient, pour ce genre de littérature, inhabituellement claires et logiques.

Un premier programme a alors été réalisé en faisant correspondre au numéro de position de chaque ancêtre un nombre, composé de 0 et de 1, qui soit son équivalent binaire. Les nombres, sur la 2200, étant limités à 13 chiffres, il n'était pas possible théoriquement de dépasser la 13^e génération. En pratique, la place occupée par cette traduction (8 octets par position d'ancêtre) permettait tout juste d'atteindre 9 générations. Les conversions étaient d'autre part relativement longues et, comme elles étaient très nombreuses, les temps d'exploitation pouvaient convenir pour des tests mais excluaient tous calculs en grande série au-delà de 5-6 générations.

Une méthode générale de calcul de consanguinité basée sur la numérotation binaire a cependant pu être ainsi mise au point. Pour réduire les temps de calcul, elle a été transposée, toujours en BASIC, sur l'I.B.M. 370 du C.I.R.C.E. Evidemment beaucoup plus rapides les exploitations ont été rapidement limitées par les prix de revient.

Les possibilités du binaire de la 2200 ont alors été à nouveau explorées et ont conduit à un programme général, allant jusqu'à 16 générations, et donnant des temps d'exploitation très acceptables avec des coûts extrêmement réduits.

La numérotation binaire des généalogies et sa traduction en mémoire

La numérotation décimale des positions des individus d'une généalogie et son équivalent binaire sont donnés dans le tableau 1, pour 6 générations. Le mode de codification pour les générations suivantes est évident. Le numéro de génération correspond au nombre de chiffres de la position en binaire.

Or, dans la 2200 comme dans la plupart des ordinateurs, les positions binaires vont par octet : deux octets suffisent donc pour mémoriser la position de tout ancêtre jusqu'à la 16^e génération.

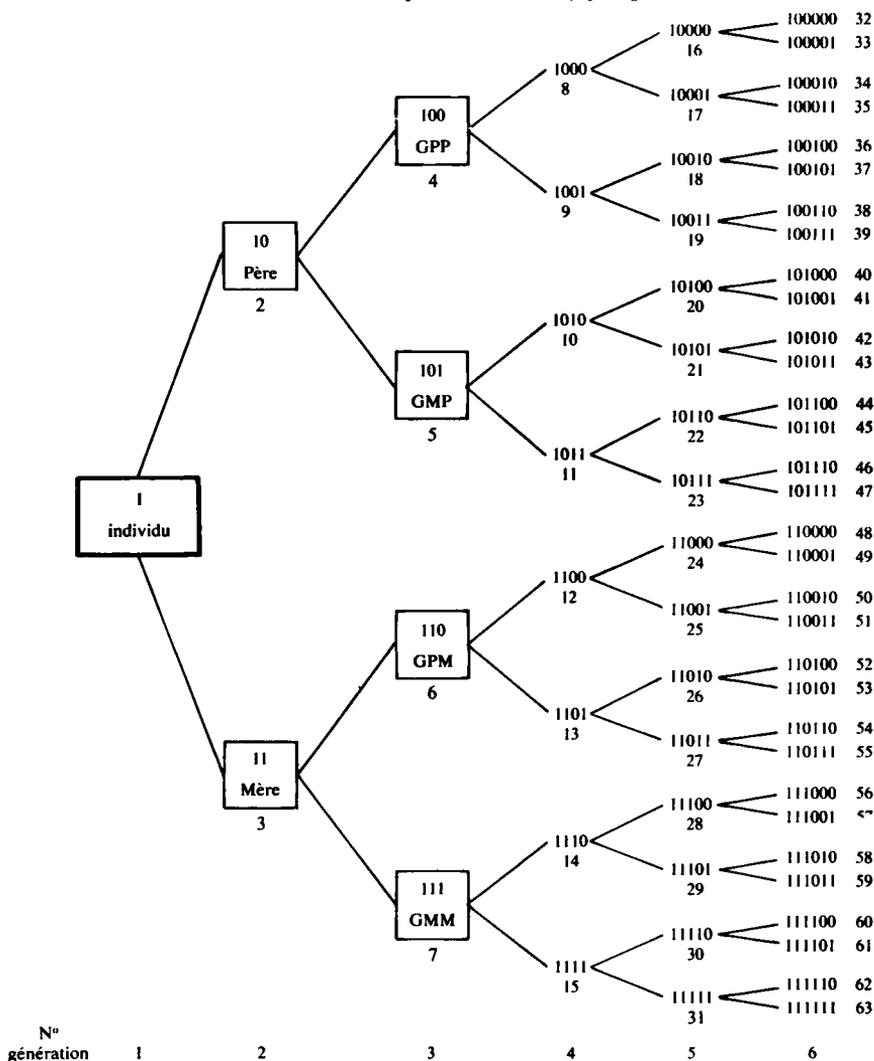
Comme nous disposions de plusieurs milliers de généalogies sur 10 générations et que les essais précédemment évoqués avaient montré que le taux de consanguinité pouvait croître de façon non négligeable de la 7^e à la 9^e, la solution d'un seul octet n'a pu être retenue. En cas de besoin, l'extension à 3 octets, soit 24 générations, ne pose aucun problème mais l'information nécessaire est rare et coûteuse à réunir.

Le principe du calcul de la consanguinité est simple. Mais en pratique, dès que le nombre des individus qui se retrouvent plus d'une fois dans l'origine dépasse quelques unités, que leurs positions se multiplient et s'éloignent, il n'est pas aisé, aussi bien en calcul manuel qu'en ordinateur :

- d'éliminer systématiquement tous les individus qui se trouvent derrière un ascendant déjà pris en considération,
- de tenir compte de la consanguinité la plus élevée d'un tel ascendant.

TABLEAU 1

Numérotations décimale et binaire d'une généalogie
Decimal and binary numerations of pedigree



Pour respecter la première condition la solution binaire est simple : tous les ascendants d'un individu ont un numéro binaire de position commençant par le numéro de cet individu. Il est facile de le vérifier sur le tableau 1 : tous les ascendants de la GMP ont un numéro commençant par 101, tous ceux de la mère débutent par 11, etc.

Recherche de toutes les voies pouvant apporter de la consanguinité

Tous les cas possibles de consanguinité sont tout d'abord détectés en prenant les individus par numéro décroissant de génération (et évidemment par sexe : 0 pour les mâles, 1 pour les femelles dans le dernier chiffre binaire du numéro de position). La numérotation binaire donne facilement la position de l'ancêtre qui reçoit une consanguinité d'un même individu situé en deux positions antérieures : son numéro est composé des premiers chiffres communs de ces deux positions.

Si par exemple le même animal est en 100000 et 100010, il apportera de la consanguinité à la position 1000. S'il est en 100011 et 10101, la position 10 (le père) sera concernée. Les positions 101000 et 111100 n'influencent que sur 1...

La détection de ces chiffres binaires communs se fait par des décalages, extrêmement rapides en binaire. Prenons le deuxième exemple. Nous avons en mémoire :

00000000 00100011 et 00000000 00010101

Après cadrage à gauche les mémoires contiennent :

10001100 00000000 et 10101000 00000000

Une comparaison booléenne met des 0 dans les positions semblables et des 1 dans celles qui diffèrent. La mémoire de gauche devient alors :

00100100 00000000

Ayant ajouté deux autres octets à gauche, il est alors facile d'y faire apparaître la partie commune en cadrant à nouveau à gauche le 3^e octet :

00000000 00000000 10010000 00000000

et en opérant des décalages simultanés sur la mémoire de droite :

00000000 00000010 10100000 00000000

Les deux premiers octets de la seconde mémoire contiennent le numéro de position du « receveur » (ici 10 : le père).

La valeur de la consanguinité reçue est déduite immédiatement du nombre des 3 séries de décalages qui ne dépendent que des numéros de génération du receveur et des deux « donneurs ».

Pour chaque valeur de consanguinité trouvée on range en mémoire sur 12 octets :

- le numéro binaire du receveur 2 octets
- les numéros des 2 positions du donneur 4 octets
- la valeur de cette consanguinité, en format .xxxxx 6 octets

Elimination des voies déjà prises

La suppression des consanguinités apportées par des ancêtres se trouvant derrière deux ascendants déjà pris en compte s'opère par groupe de voies aboutissant au même receveur. L'ensemble des receveurs ayant été triés sur leur numéro de position (2 octets) l'opération est faite séquentiellement.

Si un groupe ne contient qu'une voie elle est aussitôt imprimée. S'il en comporte plusieurs, la 1^{re} est comparée systématiquement à toutes les autres du groupe, pour l'éliminer si ses 3^e et 4^e octets d'une part, ses 5^e et 6^e d'autre part commencent par les mêmes chiffres binaires que les octets correspondants d'une voie suivante.

En reprenant le premier schéma, supposons que le même individu se trouve en position 111 (GMM) et 1101. Il y aura évidemment identité d'indicatif en 1110 et 11010, 1111 et 11011... D'après notre ordre de tri nous rencontrons en tête de groupe :

```
00000000 00000011 00000000 00110100 00000000 00111000 .xxxxx
```

et en queue

```
00000000 00000011 00000000 00001101 00000000 00000111 .xxxxx
1           2           3           4           5           6           7
```

Les 3^e et 4^e octets d'une part, les 5^e et 6^e d'autre part commencent bien par les mêmes chiffres binaires, respectivement 1101 et 111. La 1^{re} voie doit donc être éliminée. On utilise alors une démarche très semblable à celle utilisée précédemment. Les octets 3 et 4 sont transférés à droite de la mémoire auxiliaire à 4 octets :

```
00000000 00000000 00000000 00110100
```

et :

```
00000000 00000000 00000000 00001101
```

puis le cadrage à gauche du 3^e octet est opéré :

```
00000000 00000000 11010000 00000000
```

et :

```
00000000 00000000 11010000 00000000
```

Du nombre le plus grand de décalages on déduit la longueur du numéro de position le plus faible (4 dans le cas présent).

En faisant de nouveaux décalages de cette longueur dans les deux mémoires apparaissent, dans les deux premiers octets, les parties dont il n'y a plus qu'à tester l'éventuelle égalité :

00000000 00001101 00000000 00000000

00000000 00001101 00000000 00000000

S'il y a inégalité, la voie est conservée. Sinon on opère de même sur les 5^e et 6^e octets :

00000000 0011100

et

00000000 00000111

Après transfert dans les 2 mémoires auxiliaires de 4 octets et décalages nous obtenons :

00000000 00000000 11100000 00000000

et :

00000000 00000000 11100000 00000000

puis :

00000000 00000111 00000000 00000000

et :

00000000 00000111 00000000 00000000

S'il y a égalité dans les deux premiers octets (comme dans le cas présent), la voie est éliminée. Sinon elle est comparée de la même façon à toutes les autres du groupe.

On opère ainsi pour toutes les voies du groupe et on édite les consanguinités qui subsistent en les triant :

— par indicatif d'individu donneur, car les effets de la consanguinité peuvent être très différents selon qu'elle provient de tel ou tel reproducteur ;

— par nombre croissant de générations prises en compte. En effet, si le nombre d'ancêtres double à chaque génération, le nombre de voies croît généralement beaucoup plus vite. Pour une population donnée, on peut ainsi déterminer la limite au-delà de laquelle le complément d'information devient très faible par rapport à son coût.

La même démarche est reprise pour les groupes successifs, c'est-à-dire par ordre décroissant de numéro de générations du receveur.

Pour chaque indicatif d'ascendant, la valeur maximum de consanguinité reçue est mise à jour au fur et à mesure.

Amélioration de la capacité et de la vitesse d'exploitation

Comme nous l'avons signalé, la limite de 16 générations peut être portée à 24 et même au-delà par tranche de 8. Le facteur limitant est essentiellement le nombre de voies possibles. Pour pallier cet inconvénient ne sont mémorisées que les consanguinités au moins égales à 10^{-5} , plusieurs tests ayant montré que les résultats finaux sont pratiquement inchangés.

Des gains sensibles de temps et de place sont obtenus en n'introduisant que les ancêtres figurant au moins 2 fois dans l'ascendance. Le seul inconvénient est de ne pas avoir en fin de calcul les indicatifs des autres ancêtres qui ont aussi reçu de la consanguinité. Ces indicatifs peuvent cependant être retrouvés avec les numéros de position.

Les problèmes de programmation

Les difficultés de mise au point sont venues du fait que, sur la 2200, ne peuvent être affichés que les équivalents en hexadécimal des nombres binaires. De nombreux transcodages manuels ont été nécessaires.

Les positions binaires, d'autre part, ne peuvent être manipulées que par groupe de 8 — un octet — et la recherche, parmi la vingtaine d'opérations logiques disponibles, de celles permettant au mieux les décalages, extractions et comparaisons décrites précédemment, n'est pas immédiate. Mais, en définitive, les gains de vitesse et de capacité de traitement sont spectaculaires.

Application au cas classique de COMET

L'origine de ce taureau fameux est généralement fournie sur 6 générations dans la littérature (CRAPLET, 1963, p. 188).

Les ancêtres communs et leurs positions, tels qu'ils sont introduits dans la 1^{re} partie du programme, sont donnés dans le tableau 2.

La généalogie peut être imprimée, avec les numéros décimaux de position (tableau 3).

Sur les 29 voies possibles de consanguinité, 20 sont éliminées. Les 9 autres sont classées par ordre décroissant de génération (tableau 4).

TABLEAU 2

*Ascendants de Comet sur cinq générations**Ancestors of Comet for five generations*

		1	2	3	4	5
1	ALCOCK 'B	022	030			
2	BOLINGBROK	004	012			
3	FAVORITE 2	002	006			
4	FAVORITE C	011	015	019	027	
5	FOLJAMBE	008	010	014	024	026
6	HAUGHTON	017	021	029		
7	PHOENIX	005	007	013		
8	R BARKER	016	020	028		
9	YOUG PH	003				
10	YOUG STR	009	025			

27 POSITIONS

TABLEAU 3

*Origine de Comet sur cinq générations**Pedigree of Comet for five generations*

	002 FAVORITE 2	003 YOUG PH
4 BOLINGBROK	8 FOLJAMBE	16 R BARKER
	9 YOUG STR	17 HAUGHTON
		18
		19 FAVORITE C
5 PHOENIX	10 FOLJAMBE	20 R BARKER
	11 FAVORITE C	21 HAUGHTON
		22 ALCOCK 'B
		23
6 FAVORITE 2	12 BOLINGBROK	24 FOLJAMBE
	13 PHOENIX	25 YOUG STR
		26 FOLJAMBE
		27 FAVORITE C
7 PHOENIX	14 FOLJAMBE	28 R BARKER
	15 FAVORITE C	29 HAUGHTON
		30 ALCOCK 'B
		31

TABLEAU 4

Consanguinité sur cinq générations de Comet
Inbreeding over five generations of Comet's ancestors

## GENERATION ## 3			
CONSANGUINITE DE 6 FAVORITE 2	PAR FOLJAMBE	EN 24 ET 26	.12505 TOTAL .12500 TOTAL GENERAL .12500
## GENERATION ## 2			
CONSANGUINITE DE 3 YOUNG PH	PAR FOLJAMBE	EN 14 ET 24	.06255 TOTAL .06250 TOTAL GENERAL .06250
	PAR PHOENIX	EN 7 ET 13	.25005 TOTAL .25000 TOTAL GENERAL .31251
			AVEC 4 GENERATIONS .25005
CONSANGUINITE DE 2 FAVORITE 2	PAR FAVORITE C	EN 11 ET 19	.06255 TOTAL .06250 TOTAL GENERAL .06250
	PAR FOLJAMBE	EN 8 ET 10	.12505 TOTAL .12500 TOTAL GENERAL .18751
			AVEC 4 GENERATIONS .12505
## GENERATION ## 1			
CONSANGUINITE DE 1 COMET	PAR FAVORITE C	EN 15 ET 19	.01568 TOTAL .01563 TOTAL GENERAL .01563
	PAR FOLJAMBE	EN 8 ET 14	.03130 TOTAL .03125 TOTAL GENERAL .04689
	PAR PHOENIX	EN 5 ET 7	.12505 TOTAL .12500 TOTAL GENERAL .17189
	PAR FAVORITE 2	EN 2 ET 6	.29693 TOTAL .29688 TOTAL GENERAL .46877
			AVEC 3 GENERATIONS .42190
			AVEC 4 GENERATIONS .45315
CONSANGUINITE DES ASCENDANTS			
FAVORITE 2			.18751
YOUNG PH			.31251

Intérêt des exploitations sur micro-ordinateurs

Les calculs de consanguinité, même avec des engins puissants, sont des opérations longues : leur prix fait beaucoup plus que doubler généralement lorsqu'une génération supplémentaire est prise en considération. Il est trop souvent admis *a priori* qu'au-delà de 5 ou 6 générations le supplément d'information est négligeable. Or l'étude de 1 660 taureaux Normands testés a montré d'une part que la consanguinité augmente souvent de façon nette au-delà de la 7^e génération et, d'autre part, que de très faibles différences de taux de parenté, de l'ordre de 2 p. 100, pouvaient entraîner des effets importants (LEFEBVRE *et al.*, 1981).

Pour pouvoir tenir compte du maximum d'information disponible sur les origines il est donc indispensable de rechercher les moyens de traitement les plus économiques. Les très grands ordinateurs sont certainement les plus efficaces pour établir de longues généalogies à partir de gros fichiers d'état civil et extraire, de chacune d'elles, la liste des ancêtres qui s'y trouvent au moins deux fois, avec leurs numéros de position. Mais le calcul proprement dit des consanguinités et des parentés est, avec le système exposé, beaucoup plus économique sur micro-ordinateur car, s'il est 5 à 10 fois plus long, le prix de l'unité de temps y est extrêmement moins élevé.

Les connexions entre les 2 catégories de matériel étant désormais simples et très abordables, de telles solutions d'exploitation mixte se développeront rapidement.

Reçu pour publication en octobre 1981.

Summary

Use of binary numeration for calculation of inbreeding

Binary numeration of the positions of ancestors in a pedigree considerably reduces the space needed in memory banks for computer calculation of inbreeding. It was used to design a program for microcomputer which takes into account up to 16 generations. Its simplicity and low cost of implementation make possible long series of calculations.

Références bibliographiques

- ALFONSO PONCE P., 1971. Redes genealogicas y computacion de coeficientes de parentesco y consanguinidad. *Arch. Zootech.*, **20**, 257-299.
- ALFONSO PONCE P., 1974. Computacion de coeficientes de consanguinidad y de parentesco. 1^{er} Congrès mondial de Génétique appliquée à l'élevage, **2**, 449-457.
- CRAPLET C., 1953. Génétique et élevage. Vigot Frères, Paris, 292 p.
- CRUDEN D., 1949. Computation of inbreeding coefficients for closed populations. *J. Hered.*, **40**, 248-251.

- EMIK D., TERRIL C.E., 1949. Systematic procedures for calculating inbreeding coefficients. *J. Hered.*, **40**, 51-55.
- FALCO M.J., 1975. The calculation of inbreeding coefficients. *Biometrics*, **31**, 993.
- HAZEL L.N., LUSH J.L., 1950. Computing inbreeding and relationships coefficients from punched cards. *J. Hered.*, **41**, 301-306.
- KOWNACKI M., 1963. Méthodes simplifiées du calcul des coefficients d'inbreeding et de la consanguinité (en polonais). *Przeł. Hodowl.*, **31**, 33-34.
- KRAVCHENKO N.A., MAYBORODA N.N., 1968. On the methods for calculation of inbreeding coefficients and genetic affinity (en russe, rés. angl.). *Tsitol. Genet.*, **2**, 104-118.
- LEFEBVRE J., 1965. Intérêt de la numération binaire dans les recherches généalogiques. *Ann. Zootech.*, **14**, 299-305.
- LEFEBVRE J., HAMON Marie-Claire, CALLENS V., WIMITZKY M. Parentés entre 28 taureaux Normands tête de lignée et distances entre les groupes de leurs 669 fils testés (en cours de publication).
- LOWE P.C., 1972. A computer program for calculating inbreeding coefficients in Dioecious populations. U.S.D.A. (Miscellaneous publ., n° 1 220), 18 p.
- PLUM M.D., 1952. The computation of inbreeding and relationship coefficients in populations with a relatively small number of different males ancestors. *J. anim. Sci.*, **11**, 742.
- PLUM M.D., 1954. Computation of inbreeding and relationship coefficients. *J. Hered.*, **45**, 92.
- REHFELD C.E., BACUS J.W., PAGELS J., DIPERT M.H., 1967. Computer calculation of Wright's inbreeding coefficient. *J. Hered.*, **58**, 81-84.
- ROSTRON J., 1978. On the computation of inbreeding coefficients. *Ann. Hum. Genet. Lond.*, **41**, 469-475.
- WRIGHT S., 1949. Coefficients of inbreeding and relationship. *Am. Nat.*, **56**, 330-338.